Reduced-Rank Regression

Applications in Time Series

Raja Velu
rpvelu@syr.edu

Whitman School of Management
Syracuse University

Gratefully to my teachers GCR & TWA
June 6, 2008
Outline

- Brief Review of Reduced Rank Regression
- VAR Model Structure and RR Features
- Numerical Example
- Recent Developments - Forecasting/RRR-Shrinkage
Review of RRR Results

- Anderson (1951)

\[ Y_t = C X_t + D Z_t + \epsilon_t \]  \hspace{1cm} (1)

\[ m \times 1 \quad n_1 \times 1 \quad n_2 \times 1 \]

\[ \text{Rank } (C') = r < m \]  \hspace{1cm} (2)

- Two implications:

\[ l'_i C = 0 \quad i = 1, 2, \ldots (m - r) \]  \hspace{1cm} (3)

\[ C = AB \]  \hspace{1cm} (4)
Estimation of Anderson's Model

Minimize,

\[ |W| = \left| \frac{1}{T}(Y - ABX - DZ)(Y - ABX - DX)' \right| \]

\[ = \left| \tilde{\Sigma}_{\epsilon\epsilon} + (\tilde{C} - AB)\hat{\Sigma}_{xx.z}(\tilde{C} - AB)' \right| \]

(5)

\[ \tilde{\Sigma}_{\epsilon\epsilon} \] — Residual Covariance matrix under full rank

\[ \hat{\Sigma}_{xx.z} \] — Partial Covariance matrix

Thus, SVD on

\[ \tilde{\Sigma}_{\epsilon\epsilon}^{-1/2} \tilde{C} \hat{\Sigma}_{xx.z}^{1/2} \] yields Gaussian Estimates
Let $\hat{R} = \tilde{\Sigma}^{-1/2} \tilde{C} \tilde{\Sigma}_{xx,z} \tilde{C}' \tilde{\Sigma}^{-1/2}$; Let $V_j$ be a normalized eigen vector of $\hat{R}$ associated with $j^{th}$ largest eigen value, $\hat{\lambda}_j^2$

$$\hat{C} = \hat{A}\hat{B} = P\tilde{C} \tag{6}$$

$$\hat{A} = \tilde{\Sigma}^{-1/2}\hat{V}_{(r)} \quad \hat{B} = \hat{V}_{(r)}' \tilde{\Sigma}^{-1/2}\tilde{C}$$

where $\hat{V}_{(r)} = [\hat{V}_1, \ldots, \hat{V}_r]$, 

- Estimation of $l_i'$ s in (3)

$$\hat{l}_i' = \hat{V}_{(i)}' \tilde{\Sigma}^{-1/2} \quad i = r + 1, \ldots, m \tag{7}$$

- Unknown Parameters: $mn_1 \to r(m + n_1 - r)$
Review (contd.)

Note $|\hat{\Sigma}_{ee}| = |\tilde{\Sigma}_{ee}| \prod_{j=r+1}^{m} (1 + \hat{\lambda}_j^2)$ (8)

$(1 + \hat{\lambda}_j^2) = 1/(1 - \hat{\rho}_j^2)$

$\hat{\rho}_j^2$ : Squared Partial Canonical Correlations

LR Test for $H_0 : \text{Rank } (C) \leq r$

$[T - n_1 + (n_1 - m - 1)/2] \cdot \sum_{j=r+1}^{m} \log(1 - \hat{\rho}_j^2)$ (9)

$\sim \chi^2_{(m-r)(n_1-r)}$

Hence, RRR $\iff$ Canonical Correlation Analysis
VAR Models

- Number of Parameters depend on order ‘$p$’ and dimension ‘$m$’

- Stationarity imposes constraints on the AR Coefficient matrices


- Box & Tiao (1977): Cannon Corr/Non-stationarity – Cointegration ideas
VAR/RR - Features

\[ Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \epsilon_t \]  \hspace{1cm} (10)

- \( \Phi_2 \) – reduced-rank; \( \Phi_1 \) – Full rank – Anderson’s Model

- Rank \((\Phi_1 : \Phi_2)\) lower rank;

\[ Y_t = A(B_1 : B_2) \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \end{pmatrix} + \epsilon_t \]  \hspace{1cm} (11)

(Velu et al 1986)

\[ Y_t = A_1 BY_{t-1} + A_2 BY_{t-2} + \epsilon_t \]  \hspace{1cm} (12)

(Reinsel 1985) - Index Model

- Let \( \Phi_j = A_j B_j \) and range \((A_{j+1}) \subset \text{range} \( (A_j) \)

(Ahn and Reinsel, 1988) – Nested RR
VAR- Unit Root Features

- Model (10): \( Y_t - \Phi_1 Y_{t-1} - \Phi_2 Y_{t-2} = \Phi(L)Y_t = \epsilon_t \)
  \(|\Phi(L)| = 0\), has \( d \leq m \) unit roots; rest are outside unit circle

- Error Correction (Granger) form:
  \[ W_t = CY_{t-1} - \Phi_2 W_{t-1} + \epsilon_t \] (13)
  \( \text{Rank } (C) = \text{Rank } (-\Phi(1)) \leq r = m - d \) (14)

- Thus, non-stationarity can be checked via reduced-rank test; Anderson (1951)

- Implications of (14); \( W_t = Y_t - Y_{t-1} \) is stationary
VAR- Unit Root Features - Interpretation


Note (13) is: $W_t = ABY_{t-1} - \Phi_2 W_{t-1} + \epsilon_t \ldots \ldots (14)$

- $BY_t$ are stationary; rows of $B$ are co-integrating vectors

- Let $Z_{t \times 1} = \begin{pmatrix} B \\ B^* \end{pmatrix} Y_t$, $B^* (d \times m)$ is orthogonal to $B$

- $B^* Y_t$ are non-stationary – “Common Trends”
VAR-Unit Root-RRR - Advantages

- Parsimonious Parametrization
- Exhibit any “co-features” or “common features”
- Improves forecasting

Other References are:

- Tiao and Tsay (1988)
- Hannan and Deistler (1988)
- Engle and Kozicki (1993)
- Min and Tsay (2005)
Cointegration-Unit Root-Tests

- $H_0: \text{Rank } (C) = r \ \text{vs} \ H_1: \text{Rank } (C) < r$

- Johansen’s Trace Statistics:
  \[ L_{tr}(r) = -(T - mp) \sum_{j=r+1}^{m} \ln(1 - \hat{\rho}_j^2) \ldots \ldots \ldots \ldots (9) \]

- Limiting distribution of $L_{tr}$ is a standard Brownian motion, depends only on ‘r’

- Estimation of $C, A, B$ etc. all follow from Anderson (1951)
Beyond Cointegration – Common Features

- Note that in Model

\[ W_t = ABY_{t-1} - \Phi_2 W_{t-1} + \epsilon_t, \]

the component \( BY_{t-1} \) is stationary

- Find \( l_i' \ni l_i' A = 0 \) spans the co-feature space (Anderson 1951)

- Find \( l_i' \ni l_i' A = 0 \) and \( l_i' \Phi_2 = 0 \), Serial Correlation Common Feature (Engle & Kozicki 1993)

- Common Seasonality (Hecq et al 2006)

Numerical Example

- Four Quarterly Series: Private Investment, GNP, Consumption and Unemployment, 1951-1988, T = 148 (See Figure 1)

- VAR(2) is specified; also \( \text{rank} (\Phi_2) = 1 \) based on partial Canonical Correlations (See Table 2)

- Recall that Partial Canonical analysis between \( W_t \) and \( Y_{t-1} \), given \( W_{t-1} \Leftrightarrow LR \) for Unit Roots;

  Results suggests \( d = 2 \) unit roots, thus \( r = 2 \) cointegrating ranks; (See Table 3)

- Model (14): \( W_t = CY_{t-1} - \Phi_2 W_{t-1} + \epsilon_t \)
Figure 1

Fig. 1. Logs of Quarterly US Economy Time Series, 1951 - 1988

- Private Investment
- GNP
- Consumption
- Unemployment

(13 A)
### Table 2

#### SPECIFICATION OF AR ORDER AND REDUCED RANKS

Table 2. Summary of Test Statistics for AR Order $j$ and Ranks $r_j = m - s$ of Coefficients $\Phi_j$ for US Economy Data, Partial Canonical Correlation Analysis between $Y_t$ and $Y_{t-j}$, given $Y_{t-1}, \ldots$

<table>
<thead>
<tr>
<th>$j$</th>
<th>Eigenvalues</th>
<th>$C(j, s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$s=4$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(r=0)$</td>
</tr>
<tr>
<td>1</td>
<td>0.999, 0.931, 0.825, 0.540</td>
<td>1895.95</td>
</tr>
<tr>
<td>2</td>
<td>0.374, 0.057, 0.004, 0.001</td>
<td>72.69</td>
</tr>
<tr>
<td>3</td>
<td>0.098, 0.051, 0.027, 0.002</td>
<td>24.23</td>
</tr>
<tr>
<td>4</td>
<td>0.107, 0.071, 0.011, 0.001</td>
<td>25.34</td>
</tr>
<tr>
<td>5</td>
<td>0.146, 0.033, 0.012, 0.002</td>
<td>25.02</td>
</tr>
<tr>
<td>6</td>
<td>0.102, 0.064, 0.012, 0.004</td>
<td>22.10</td>
</tr>
</tbody>
</table>

- Eigenvalues are squared partial canonical correlations $\hat{\rho}_i^2(j)$. Test for
  
  $H_0: \text{rank}(\Phi_j) = m - s$ is $C(j, s) = -(T - j - mj - 1) \sum_{i=(m-s)+1}^{m} \log[1 - \hat{\rho}_i^2(j)] \sim \chi^2_{s_2}$
Table 3

SPECIFICATION OF \( \text{rank}(C) \) OR NUMBER \( d \) OF UNIT ROOTS

Table 3. Summary of LR Tests for \( \text{rank}(C) = r \) (cointegration) or number of unit roots \( d = m - r \) for US Economy Data, Partial Canonical Correlation Analysis between \( W_t \) and \( Y_{t-1} \), given \( W_{t-1}, \ldots \).

<table>
<thead>
<tr>
<th>( j )</th>
<th>Eigenvalues</th>
<th>( r=0 ) ((d=4))</th>
<th>( r=1 ) ((d=3))</th>
<th>( r=2 ) ((d=2))</th>
<th>( r=3 ) ((d=1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.304, 0.115, 0.060, 0.011</td>
<td>81.20</td>
<td>28.36</td>
<td>10.56</td>
<td>1.56</td>
</tr>
<tr>
<td>2</td>
<td>0.207, 0.086, 0.053, 0.012</td>
<td>56.43</td>
<td>22.81</td>
<td>9.72</td>
<td>1.78</td>
</tr>
<tr>
<td>3</td>
<td>0.196, 0.085, 0.074, 0.010</td>
<td>56.60</td>
<td>25.24</td>
<td>12.52</td>
<td>1.41</td>
</tr>
<tr>
<td>4</td>
<td>0.219, 0.078, 0.072, 0.011</td>
<td>59.20</td>
<td>23.93</td>
<td>12.26</td>
<td>1.59</td>
</tr>
<tr>
<td>5</td>
<td>0.179, 0.089, 0.066, 0.013</td>
<td>52.80</td>
<td>24.84</td>
<td>11.57</td>
<td>1.83</td>
</tr>
<tr>
<td>6</td>
<td>0.196, 0.099, 0.068, 0.008</td>
<td>56.50</td>
<td>25.75</td>
<td>11.11</td>
<td>1.17</td>
</tr>
</tbody>
</table>

- Eigenvalues are squared partial canonical correlations \( \hat{\rho}_i^2 \) of \( W_t \) and \( Y_{t-1} \),
- \( -(T - j) \log(\lambda) = -(T - mp) \sum_{i=r+1}^{m} \log(1 - \hat{\rho}_i^2) \) is LR statistic.
Estimation with ‘Cointegrating’ Reduced Rank Imposed

\[ \hat{C} = \hat{A} \hat{B} = \begin{bmatrix} -0.241 & 0 \\ -0.022 & 0 \\ -0.01 & 0.107 \\ -0.209 & -0.921 \end{bmatrix} \begin{bmatrix} 1 & 0 & -0.997 & 0.092 \\ 0 & 1 & -0.942 & 0.094 \end{bmatrix} \]

\[ = \begin{bmatrix} -0.241 & 0 & 0.24 & -0.222 \\ -0.022 & 0 & 0.022 & -0.002 \\ -0.01 & 0.107 & -0.09 & 0.009 \\ -0.209 & -0.921 & 1.07 & -0.105 \end{bmatrix} \]

\[ -\hat{\Phi}_2 = \begin{bmatrix} 1 \\ 0.167 \\ 0.070 \\ -1.673 \end{bmatrix} \begin{bmatrix} 0.239 & 0 & 2.724 & 0 \\ 0.04 & 0 & 0.45491 & 0 \\ 0.017 & 0 & 0.19068 & 0 \\ -0.400 & 0 & -4.5573 & 0 \end{bmatrix} \]

\[ \hat{\Sigma}_\epsilon = 4.503 \times 10^{-15} \text{ - Estimates } \sim \text{ Full Rank LSES} \]
Numerical Example (contd.)

- Implications / Interpretation

- Cointegration: Stationary Combinations:
  \[ Z_{3t}^* = Y_{1t} - 0.997Y_{3t} + 0.042Y_{4t} \]
  \[ Z_{4t}^* = Y_{2t} - 0.942Y_{3t} + 0.094Y_{4t} \]

- Common Trend Components:
  \[ Z_{1t}^* = b_1^t Y_t \text{ and } Z_{2t}^* = b_2^t Y_t \]
  where \( b_1^t A \) and \( b_2^t A = 0; \)
  choose \( b_1 \ni b_1^t \Phi_2 = b_1^t a_3 d^t = 0 \)

Thus, \( b_1^t W_t = b_1^t \xi_t \), so
\[ Z_{1t}^* = b_1^t Y_t \text{ is a random walk} \]
\[ Z_{1t}^* = -0.132Y_{1t} + Y_{2t} + 0.285Y_{3t} + 0.033Y_{4t} \]

See Figure 2
Figure 2

Fig. 2. Cointegrating & Common Trend Series for US Economy Data, 1951 - 1988

Cointegrated Series, z3

Cointegrated Series, z4

Common Trend Series, z1

Common Trend Series, z2
Numerical Example (contd.)

- Overall interpretation;

\[ Z_t = \left( \frac{B}{B^*} \right) Y_t = \left( \frac{Z_{1t}}{Z_{2t}} \right) \]

\[ Z_{1t} - Z_{1t-1} = (B^* a_3) d' W_{t-1} + \epsilon_{1t}^* \]

\[ Z_{2t} = (I + BA) Z_{2t-1} + (Ba_3) d' W_{t-1} + \epsilon_{2t}^* \]

\[ B^* a_3 = \begin{bmatrix} 0 \\ -.289 \end{bmatrix}, \quad Ba_3 = \begin{bmatrix} 0.776 \\ -.055 \end{bmatrix}, \]

\[ I + BA = \begin{bmatrix} .750 & -.191 \\ -.032 & .814 \end{bmatrix} \]

\[ Z_{1t} \sim RW; \quad Z_{2t}^* - Z_{2t-1}^* \sim \text{with AR(1) structure}, \quad Z_{3t}^* \sim \text{AR(2)}; \]

\[ Z_{4t}^* \sim \text{AR(1)} \]
Effect of RR estimation on prediction, \( \text{Rank} \ (C) = 1; \ C = \alpha \beta' \)

Covariance matrix of the prediction error:

\[
(1 + x'_0 (XX')^{-1} x_0) \Sigma_{\epsilon \epsilon} - \\
\left\{ x'_0 (XX')^{-1} x_0 - \frac{(\beta' x_0)^2}{\beta' (XX')^{-1} \beta} \right\} \left( \Sigma_{\epsilon \epsilon} - \Sigma_{\epsilon \epsilon}^{1/2} V_1 V_1' \Sigma_{\epsilon \epsilon}^{1/2} \right)
\]

(Reinsel and Velu, 1998, p46)

Empirical Analysis of Hog data indicate accounting for both non-stationarity and rank reduction consistently outperforms well at all forecasting horizons (Wang and Bessler 2004)
RRR and Shrinkage

Recall

$\hat{C} = \hat{A}\hat{B} = \hat{\Sigma}_{yy}^{1/2}\hat{V}(r)^{\prime}\hat{V}(r)\hat{\Sigma}_{yy}^{-1/2}\tilde{C}$

$= \hat{H}_{1}^{-1}\hat{\Delta}_{1}\hat{H}_{1}\tilde{C}$

where $\hat{\Delta}_{1} = Diag(1, 1, \ldots, 1, 0\ldots0), m \times m$ diagonal matrix.

GCV - Breiman and Friedman (1997)
Reinsel (1998)

$\hat{\Delta}_{1} = Diag(\hat{d}_{j}), \hat{d}_{j} = \frac{(1-p_{1})(\hat{\rho}^{2}_{j}-p_{1})}{(1-p_{1})^{2}\hat{\rho}^{2}_{j} + p_{1}^{2}(1-\hat{\rho}^{2}_{j})} \quad j = 1\ldots m$

where $p_{1} = n_{1}/(T - n_{2})$
Forecasting Large Datasets (Carriero, Kapetanoos & Marcellino 2007)

- RRR (Reinsel and Velu, 1998)
- BVAR (Doan et al, 1984)
- BRR (Geweke, 1996)
- RRP (Restrictions on the posterior mean)
- MB (Brieman, 1996)
- FM (Stock and Watson 2002)

Data: 52 US Macroeconomic series (Real, Monetary & Financial)

Three Benchmarks: AR (1), BVAR and RW;

Among the six models: BRR - Short horizons
RRP - Long Horizons
FM - Best for 1-step ahead.

Conclusion: Using Shrinkage and Rank reduction in combination
Future Work

- Shrinkage - RRR, Needs further investigation, Yuan et al 2007- RRR/ LASSO


- More broadly, in VAR context, the impact of misspecification of both the order ‘p’ and rank ‘r’ need to be studied.